AFFTC-PA-05069

UNMANNED AIR VEHICLES: A NEW AGE IN HUMAN FACTORS EVALUATIONS

John J. Spravak
Deborah A. Moisio
Mary G. Payton

AIR FORCE FLIGHT TEST CENTER
EDWARDS AFB, CA

AFFTC

AIR FORCE FLIGHT TEST CENTER
EDWARDS AIR FORCE BASE, CALIFORNIA
AIR FORCE MATERIEL COMMAND
UNITED STATES AIR FORCE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
May 2005	Technical Paper	September 2004 to 10 April 2005

4. TITLE AND SUBTITLE
Unmanned Air Vehicles: A New Age in Human Factors Evaluations

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
John J. Spravak
Deborah A. Moisio
Mary G. Payton

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Flight Test Center
Edwards AFB CA 93524-6001

8. PERFORMING ORGANIZATION REPORT NUMBER
AFFTC-PA-05069

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
412th Test Wing Human-Systems
Bldg 1609
30 N. Wolfe Rd.
Edwards, CA 93524-6843

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
N/A

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release: distribution is unlimited.

13. SUPPLEMENTARY NOTES
CA: Air Force Flight Test Center Edwards AFB CA CC: 012100

14. ABSTRACT
As the role of the aircraft pilot transitioned from a nearly total manual controller in early manned aircraft to one of supervisory control and/or cooperative functioning in "unmanned" aircraft, the human factors flight test approach and the associated test methodologies have necessarily changed. Piloting air vehicles evolved from using cockpit instruments and manual controls to fly the aircraft, to monitoring the cockpit instruments, which fly the aircraft nearly automatically, to using ground station instruments to fly the aircraft remotely. While most, if not all, of the physical stressors of the cockpit are absent from the typical ground control station, many of the cockpit stimuli that provide invaluable aircraft health and status information are also absent. Increased levels of automation have induced new types of failures. These include failure to monitor, vigilance decrement, over reliance on standard values, automation-induced complacency, and increased latency in detecting problems. Consequently, these failures often lead to reduced operator performance due to information shortfall.
As the pilot-aircraft relationship evolve, the focus of human factors evaluations moves from what the pilot physically perceives and processes in the cockpit, to what the pilot mentally perceives and processes on the ground. Physical information from the aircraft, such as vibration and sound cues, must be transformed into usable information on the ground station displays. This includes keeping the level of automation appropriate so the pilot on the ground can be aware of and adequately handle emergency situations. Since pilot workload for an Unmanned Air Vehicle (UAV) is mostly mental, maintaining situation awareness is paramount.
This paper identifies some critical components of new human factors approaches for evaluating UAV human-system interfaces and compares them with approaches traditionally used to evaluate manned air vehicles.

15. SUBJECT TERMS
Unmanned Aerial Vehicles, human factors, human-system integration, flight test, operator workstation, pilot-vehicle interface, evaluation

16. SECURITY CLASSIFICATION OF: UNCLASSIFIED			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
					412 TW/EN (Tech Pubs)
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(include area code)*
Unclassified	Unclassified	Unclassified	None	16	661-275-9581

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. 239.18

UNMANNED AIR VEHICLES: A NEW AGE IN HUMAN FACTORS EVALUATIONS

John J. Spravka

Deborah A. Moisio

Mary G. Payton

412 TW/ENFH
118 N. Wolfe Rd.
Edwards, CA 93524
Unites States of America

Email: john.Spravka@edwards.af.mil / deborah.moisio@edwards.af.mil / mary.payton@edwards.af.mil

SUMMARY

As the role of the aircraft pilot transitioned from a nearly total manual controller in early manned aircraft to one of supervisory control and/or cooperative functioning in "unmanned" aircraft, the human factors flight test approach and the associated test methodologies have necessarily changed. Piloting air vehicles evolved from using cockpit instruments and manual controls to fly the aircraft, to monitoring the cockpit instruments which fly the aircraft nearly automatically, to using ground station instruments to fly the aircraft remotely. While most, if not all, of the physical stressors of the cockpit are absent from the typical ground control station, many of the cockpit stimuli that provide invaluable aircraft health and status information are also absent. Increased levels of automation have induced new types of failures. These include failure to monitor, vigilance decrement, over reliance on standard values, automation-induced complacency, and increased latency in detecting problems. Consequently, these failures often lead to reduced operator performance due to information shortfall.

As the pilot-aircraft relationship evolve, the focus of human factors evaluations moves from what the pilot physically perceives and processes in the cockpit, to what the pilot mentally perceives and processes on the ground. Physical information from the aircraft, such as vibration and sound cues, must be transformed into usable information on the ground station displays. This includes keeping the level of automation appropriate so the pilot on the ground can be aware of and adequately handle emergency situations. Since pilot workload for an Unmanned Air Vehicle (UAV) is mostly mental, maintaining situation awareness is paramount.

This paper identifies some critical components of new human factors approaches for evaluating UAV human-system interfaces and compares them with approaches traditionally used to evaluate manned air vehicles.

1.0 INTRODUCTION

While the term **Unmanned** Air Vehicle (UAV) implies there is little or no human-system integration required, all UAVs require human-system interface (HIS) in both the control and maintenance of the system. Because the UAV operator is located in a ground segment, far from the aircraft itself, the methods and measures used can be very different than in a manned aircraft. Also, the differences in the environment, the operation, and crew complement require new or modified test methods and measures. This paper discusses operator HIS

only, since maintenance of a manned vehicle is not much different than that of an unmanned vehicle.

2.0 MANNED AIRCRAFT HUMAN-SYSTEM INTEGRATION EVALUATIONS

Human-system integration evaluations of manned aircraft include analyses and tests of the cockpit displays, controls, environment and ergonomics, system communications, overall task allocation, operator situation awareness, operator workload, life support systems, and personnel training. Table 1 lists these areas of investigation and some of their more specific areas of evaluation. Two types of measures of performance are used to determine how well the human has been integrated into the aircraft: subjective measures, such as how difficult a task is perceived; and objective measures, such as physiological data. Although objective measures are preferred, the use of objective measures is often not possible due to a number of factors, including limited time and funding, and technological constraints.

Table 1: Specific Human-System Integration Evaluation Areas

General Evaluation Area	Specific Evaluation Area
Cockpit Displays – Information Readability	Text character or symbol font, size, stroke width, brightness, contrast, color, and information location, organization, and formatting
Cockpit Displays – Information Interpretability	Information (text or symbology) color coding, semantic accuracy/intuitiveness, salience, clutter, density
Cockpit Environment	Life support system, vibration/noise, cockpit temperature/humidity/lighting
Cockpit Ergonomics	Cockpit external (canopy) and internal (displays) visibility, pilot reach, body clearance, and seat restraints
Controls Operability	Controls location, labelling, discriminability between co-located controls, ease-of-use, anthropometric accommodation
System Communications	Speech intelligibility
Situation Awareness	Information perception, comprehension and projection
Workload	Task type, complexity, duration, difficulty, and resource type and demand

2.1 Subjective Measures of Performance

Table 2 presents some subjective HSI measure types, examples of the areas evaluated and the types of data produced.

Table 2: Subjective Human-System Integration Measures of Performance.

Measure Type	Systems/Construct Evaluated	Measure of Performance
Subjective Rating Scales	All	System adequacy/usability
	Visual Displays	Information readability, visibility, interpretability, formatting, organization, clutter
	Aural Displays/Comm System	Readability & Strength
	Situation Awareness	Object/text perception, comprehension and projection
	Workload	Task difficulty, duration, and physiological, cognitive and emotional stress
Expert Observation	Situation Awareness, Workload	

Subjective questionnaires utilize a variety of numeric rating scales and specific or open-ended questions to gather HSI test data. Several system interface subjective rating scales have been developed over the years, all of which have their relative strengths and weaknesses. Some of the more popular rating scales are shown in Table 3.

Table 3: Subjective Rating Scales

Rating Scale	Rating Scale Type	Strengths/Weaknesses
Air Force Flight Test Center (AFFTC)	System Adequacy, 6 point., Interval, Bi-Polar	Strengths: No middle point-forced to make choice Weaknesses: Not used outside AFFTC
AFFTC-modified USAF-SAM (School of Aerospace Medicine)	Workload, 7 point., Interval	Strengths: Easy to use, fits on flight cards Weaknesses: General workload not specific
Readability & Strength	Comm Quality, 5 point., Interval	Strengths: Pilot friendly/familiar Weaknesses: Verbal anchors not defined
Bedford	Workload, Ordinal	Strengths: Pilot friendly/familiar, validated [1] Weaknesses: Mid-range semantic descriptors are vague, interchangeable
Subjective Workload Assessment Technique (SWAT)	Workload, 100 point., Interval, 3 Dimensional	Strengths: Easy to use, once learned Weaknesses: Requires card sort, specialized software, and training
Modified Cooper-Harper	Workload, Ordinal	Strengths: Pilot friendly/familiar Weaknesses: Non-interval scale
Situation Awareness Global Assessment Technique (SAGAT)	Situation Awareness	Strengths: Widely used and accepted Weaknesses: Limited to simulations

2.2 Objective Measures of Performance

While subjective measures are the norm, objective measures are the most desirable because they are direct measures of system performance. Additionally, they provide increased construct validity over that of subjective measures (i.e., does the data actually tell you what you want to know about the interface?). Table 4 lists several objective measures, the systems or constructs they are used to assess, and examples of these measures of performance.

Table 4: Objective Measures of Performance

Objective Measure	Systems/Construct Evaluated	Measure of Performance
Physiological	Workload, Stress, Cognitive Attention	Heart rate, respiration rate, core temperature, galvanic skin response, eye-blink rate, EEG, EKG
Anthropometric	Pilot Cockpit Accommodation	Design eye height, functional arm reach, leg clearance, hand/arm/leg/torso/head length/width/circumference
Perceptual	Visual Displays	Ambient light, display brightness/contrast, color wavelength/contrast
	Aural Displays & Communication System	Ambient noise, speech intelligibility
	Pilot Protection	Ambient noise, cockpit vibration, cockpit temperature, breathing air content/pressure
Operator/System Performance	Pilot Flight Controls	Aircraft altitude, attitude and airspeed error
	Visual Displays	Target detection and identification; tracking error; information processing rate/ efficiency/ accuracy
	Aural Displays/Comm System	Word recognition rate (Modified Rhyme Test)
	Situation Awareness	Situation Awareness Global Assessment Technique (SAGAT) Display object appearance detection and object status change recognition rates
	Workload	Successful task accomplishment rate, flight control performance (Altitude, Attitude, Air Speed), time on task, number of radar target lock-ups, probability of Kill (PK), operator errors

3.0 CURRENT UAV HUMAN-SYSTEM INTEGRATION EVALUATIONS

3.1 Global Hawk Human-system Interface (HIS) Characterization

During the Global Hawk Advanced Concept Technology Demonstration in 1998, the Air Force Flight Test Center (AFFTC) conducted a comprehensive evaluation of the Global Hawk UAV ground control station at Edwards Air Force Base, California [2]. Ground Segment workstation operators rated the environment, ergonomics, controls and displays, situation awareness, display windows, task procedures and task workload for specific operations and phases of flight. The operations and phases of flight examined included: taxi,

takeoff, autonomous and operator-effected navigation; communications; image collection and dissemination; execution of operator interventions; and landing.

Most of the measures of performance that are used to conduct HSI evaluations of manned aircraft are also used to evaluate unmanned systems. Table 5 lists a variety of human performance measures that have been used to assess the effect of various types of controls and displays on operators in UAV workstation simulations.

Table 5: Examples of UAV Operator Measures of Performance

Reference	Independent Variable	Experimental Conditions	Measures of Performance/ Dependent Variable
Geddes et al. [3]	Operator Control Level	Manual vs. Supervisory	Number of Operator Actions
			Sternberg Secondary Task (Cognitive Workload)
			Time on Task Flight Path Tracking
Dixon et al. [4]	Display Type	Visual vs. Auditory	Command Response Time
	Operator Control Level	Manual vs. Supervisory	Command Report Accuracy
			Command Recall Button Frequency of Use
			Target of Opportunity Detection Rate
			System Failure Detection Rate and Time
			Task Completion Time
			Inter-stimulus Interval between Serial Tasks
Calhoun et al. [5]	Sensory Input Type	Active Tactile vs. Salient Visual and/or Auditory Alerts	Task Completion Time
			Alert Reaction Time
			Airspeed, Altitude, and Flight Path Error
Draper et al. [6]	Command Input Mode	Speech vs. Manual	Frequency of Incorrect Task Completion
			Alert Reaction Time
			Airspeed, Altitude, and Flight Path Error
			Correct Speech Recognition Percentage

Table 5: Examples of UAV Operator Measures of Performance (Concluded)

Reference	Independent Variable	Experimental Conditions	Measures of Performance/ Dependent Variable
Bell and Cooke [7]	UAV Single vs. Team Operator Performance	Cognitive Ability and Individual vs. Team Performance	Time Spent in an Alarm State
			Amount of Fuel Used
			Amount of Film Used
			Number of Missed Targets
			Number of Critical Waypoints Missed
			Time Spent in a Warning State
			Route Sequence Violations
van Breda [8]	Cognitive Load Effects	Display Update Rate and Auditory Cognitive Load	Target Viewable Time Percentage
			Target Tracking Error (angular difference between cross hairs and target)
			Percentage Correct Target-Letter Hits
			Number of 3 Target-Letter Hits
Chadwick et al. [9]	Display Number	Single vs. Dual	Video/Audio Recording
	Task Switching	Infrequent - Frequent	NASA Time Load Index
	Multi-robot Task Completion Strategy	Parallel vs. Sequential	Situation Awareness Global Assessment Technique
			Task Completion Time
			Robot Casualties
			Task Switching Frequency
			Map Drawing Accuracy
			Tracking Error
			Number of Delivered Boxes

What follows is a short description of the research in which each of the measures listed in Table 5, above, were used so that the reader can 1) see the types of UAV HSI measures used in representative research and how they were used to demonstrate the associated research hypotheses, and 2) learn about some of the operator workstation design factors that affect UAV operator performance.

3.2 Supervisory vs. Manual Control

In a study designed to compare operator performance in a supervisory versus manual UAV control situation, Geddes et al. assessed the controller task demand of a small-scale uninhabited tactical aircraft (UTA) by measuring the number of operator actions required to complete a task, cognitive workload (Sternberg secondary task performance) and time on task [3]. The UTA operator control task performed under three types of operator control. The first was direct manual control where the operator issued flight control commands such as pitch, roll, thrust. The second was command control (supervisory) where the operator issued auto-pilot commands such as set altitude, set airspeed, set heading. The third type of control was task-level control

where the operator issued task-level commands (supervisory) such as 'line formation' and 'trail formation.' The command control and task control task decreased significantly over that of the direct manual control condition. The cognitive workload for the command and task control conditions also decreased. Clearly, the supervisory control tasks were characterized by much fewer task completion actions and much less cognitive load than the manual control tasks.

3.3 Auditory vs. Visual Displays and Manual vs. Supervisory Control

In a study comparing model predictions to experimental data in a multiple-UAV flight control task, 36 licensed pilots performed simulated military surveillance missions with 1 and/or 2 UAVs [4]. Pilots were responsible for navigating each UAV through a series of mission legs in one of the following conditions: 1) a baseline condition with all manual flight controls and visual displays; 2) an auditory off-load condition that provided auto-alerts and other relevant information to the auditory channel; and 3) an automation condition that provided auto-pilot control of the UAV. Pilots were responsible for mission completion, target search, and systems monitoring.

Measures of performance included flight path tracking, command response time and report accuracy, command recall button frequency of use, target of opportunity detection rate, system failure detection rate and time, task completion time, and inter-stimulus interval between serial tasks. Results indicated that the two types of task off-loads (auditory & automation) were beneficial in reducing task interference and overall workload.

3.4 Tactile vs. Visual/Auditory Alerts

Calhoun et al. examined the utility of active tactile alerts versus salient visual and/or auditory alerts in an UAV ground control station simulation [5]. Measures of performance included reaction time between alert onset and confirmation response to the alert, total task completion time and accuracy for each data input task as well as the root-mean-squared (RMS) error of airspeed, altitude, and flight path. Subjective ratings and comments were also obtained with debriefing questionnaires. The flight performance measure succeeded in showing an expected significant effect of mission difficulty, with lower performance found in the High Difficulty Level condition.

3.5 Speech Input vs. Manual Input

Draper et al. examined the utility of conventional manual input versus speech input for tasks performed by operators of a UAV control station simulator at two levels of mission difficulty [6]. Pilots performed a continuous flight navigation control task while completing eight different data entry tasks with each input modality. Operators were required to perform a continuous flight/navigation control task while responding to intermittent data entry tasks. Operator accuracy measures included frequency of task completion *"time-outs,"* frequency of tasks completed incorrectly, and percentage of speech commands correctly recognized. Response time between alert onset and confirmation response was also recorded; tasks where the alert was missed were discarded from the data pool. Root-mean-squared error of airspeed, altitude, and flight path were used to measure flight/navigation performance. Subjective ratings were obtained with debriefing questionnaires, including the Modified Cooper Harper rating scale [10]. Results showed that speech input was significantly better than manual input in terms of task completion time, task accuracy, flight/navigation measures, and pilot ratings. Across tasks, data entry time was reduced by approximately 40 percent with speech input.

3.6 UAV Team Performance

Bell and Cooke examined the relationship between two cognitive ability measures, grade point average (GPA) and verbal working memory capacity, and performance on a team UAV reconnaissance task [7]. Two experiments required three team members to maneuver a simulated UAV to take reconnaissance photos. Each of the team members assumed a different role with unique responsibilities. Low workload missions required that participants take 9 photos of various targets, whereas high workload missions required 20 photos and involved additional route constraints. The high workload manipulation produced significant reductions in team performance and in the performance of each of the three roles.

Team performance was measured using a composite score based on the result of mission variables including time each individual spent in an alarm state, amount of fuel used, amount of film used, number of missed targets, number of critical waypoints missed, time spent in a warning state, and route sequence violations. Results indicated that working memory capacity was more highly correlated with role performance and GPA was more highly correlated with team performance.

3.7 Cognitive Load Effects on Manual Tracking

In research addressing Wickens' multiple resource theory, van Breda measured operator tracking performance on a maritime unmanned air vehicle (MUAV) flight control task as a function of target visual image update rate and auditory cognitive load. Auditory cognitive load was manipulated by having the operators execute a continuous memory task (CMT) [8, 11-13]. The operator had to remember the frequency of occurrence of specific target letters verbally communicated to them via auditory headsets. There were three levels of cognitive load represented by three differing numbers of target letters. Tracking difficulty was manipulated by varying the visual target image update rate. As target image update rate decreased, tracking performance significantly increased [14]. In a task very similar to that of a Global Hawk command and control operator, subjects had to maintain a pre-specified distance between the MUAV and a target while performing the continuous memory task. Measures of performance were the percentage of time the target was viewable on the visual display, target tracking error (angular difference between cross hairs and target) and percentage correct target-letter hits and number of three target-letter hits. Results showed that tracking error significantly increased as a function of decreasing target image update rate. High verbal/cognitive workload, as measured by the continuous memory task, resulted in significantly higher CMT scores than the moderate workload level. However, CMT performance was unaffected by target image update rate.

3.8 Perceptual and Cognitive Factors Affecting Operator Control of Multiple Robots

Chadwick et al. used commercially-available video games played over local area networks to identify perceptual and cognitive problems for single operators controlling multiple, semi-autonomous entities (robots) [9]. They found that 1) True parallel operation of two robots was rarely, if ever, achieved, 2) The optimum level of task switching was characterized by a frequency that reduced operator situation awareness (SA) recovery time yet kept pace with the demands of each task, 3) In the single display condition, attentional tunnelling caused operators to lose track of robot #1 while interacting with robot #2, whereas in the dual display condition operators were able to track robot #1 with peripheral vision and accomplish robot #2's task successfully, 4) National Aeronautics and Space Administration (NASA) Time Load Index (TLX) workload scores revealed that dual display operator frustration decreased but mental demand, temporal demand and effort all increased compared to the single display condition, 5) Participants made frequent mode errors while viewing one robot and operating another, 6) Identification failures were prominent as participants did not recognize, on their own, that they were using one entity under their control to attack a second entity under their control, 7) In the single display scenario, with supervisory control over semi-autonomous robots, the

global map tool was difficult for novices to use, but less difficult for experienced players, 8) With regular direct or peripheral monitoring of the global map, task switching cues could be easily identified, 9) In general, single computer control functions were difficult to master, 10) Execution of the Situation Awareness Global Assessment Technique (SAGAT) structured interview during ongoing scenarios appeared to influence operator behavior after the SAGAT breaks and, therefore, was dropped from the study [15].

4.0 DIFFERENCES BETWEEN MANNED AND UNMANNED HUMAN-SYSTEM INTEGRATION EVALUATIONS

4.1 Evaluation Area Emphasis

The HSI evaluation areas for unmanned aircraft do not differ considerably from those applied in manned aircraft evaluations. Rather, the differences lie in the amount of emphasis placed on these evaluation areas. Two UAV HSI evaluation areas which are investigated more thoroughly are situation awareness and UAV operator control level.

4.1.1 Situation Awareness

The UAV operator, remotely located in a typical UAV ground control station, does not receive the same quality and quantity of information or feedback that a manned aircraft pilot does. UAV operators do not receive, or receive reduced amounts of, 'seat-of-the-pants' aircraft handling cues, such as 'out-of-the-window' visual weather, the smell of smoke, oil, or hydraulic fluid, and other sensory information. Therefore, the situation awareness of the UAV operator must be evaluated thoroughly. One must determine the adequacy of the UAV operator's mental model that has been induced by the information presented on the displays, the controls used to operate the UAV and the functional interaction of those controls, displays and operator. Additionally, because workload is ultimately and intimately affected by situation awareness, workload must also be evaluated in a thorough manner and the relationship between the two should be identified.

4.1.2 Operator-UAV Control Level

Although manned cockpits are becoming increasingly automated, the level of system automation present in most UAVs far exceeds that of manned cockpits. Therefore, the degree of automation and the level and type of control the operator has over the UAV must be identified and examined to ensure that both the workstation and the operator tasks are designed to accommodate the proper level of control.

Prior research in human performance has shown that the introduction of automation can be both beneficial and costly from a human-system performance perspective [16-21]. Consequences of these costs have ranged from temporary confusion [22] to the loss of human life [23-25]. Potential contributing factors include increases in mental workload, a loss of situation awareness, and skill degradation [26-30]. Operator trust and acceptance and automation-induced complacency have also been identified as potential problematic factors [31-36]. Automation-induced complacency has been identified as a contributing factor in many aviation accidents [37]. Additionally, complacency has been included as a behavioral coding category used by the Aviation Safety Reporting System (ASRS) to classify aviation incidents and accidents [38]. Parasuraman et al. found that when automation of varying reliability was used consistently (every trial) to augment human performance on a multi-task operation, operator performance was significantly reduced from 72percent manual detection of all engine malfunctions versus 37 percent and 28 percent detection for the low and high automation reliability rates, respectively [28]. Parasuraman et al. were also able to replicate Thackray and Touchstone's results

showing that automation-induced complacency did not affect single task performance [39-40]. Subsequent research showed that automation-induced complacency decreased overall system performance [40-43].

Automation can occur at differing stages of an operator task. Parasuraman, et al. identified four stages of human information processing as seen in Figure 1 below [29]. Examples of stage one elements include raw sensory data from the external environment and items that drive selective attention mechanisms. An example of stage one automation would be automatic target recognition software.

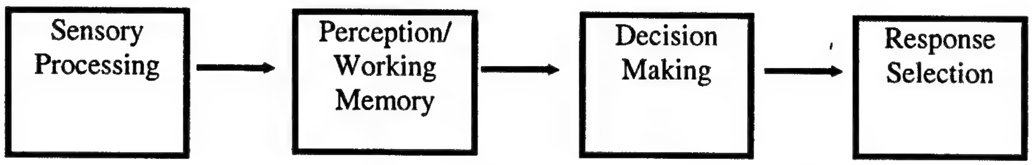

Figure 1: Four Stages of Human Information Processing.

Full conscious perception and manipulation of data in working memory (e.g., rehearsal, integration, etc.) occur in the second stage [44]. Examples of automation at this stage include the use of integrated displays, systems that present information from multiple modalities, and tools that facilitate the analysis and presentation of data. The third stage is the decision that is made based on the veracity of the data processed in stages one and two. The fourth stage is the action to carry out the decision made.

It has been found that automation may negatively affect operator performance more in the "decision making" stage of an operator's task than during the "information analysis" (perception/working memory) stage [29, 45-48]. In a series of studies, Galster set out to examine human interaction with automation in the context of a multidimensional (stage) concept of automation, as specified in the Parasuraman et al. model [29, 49-51]. Initially, Galster found that perfectly reliable automation increased human target detection performance more in the "information analysis" stage than in the "decision making" stage [49]. Contrary to previous studies, Galster also found that unreliable automation more negatively affected the information analysis stage than the decision making stage. A fourth study examined all possible combinations of manual and automated aiding in an air-to-ground search and destroy mission conducted in a high fidelity combat simulator. In the fourth study Galster found that automation in one stage positively influenced performance in subsequent stages and throughout the entire mission. These benefits were apparent in the primary task performance and the subjective ratings of mental workload (NASA TLX), situation awareness and trust in automation.

As can be seen above, the effects of aircraft automation on UAV operator performance have been well documented. However, nowhere is the phenomenon of automation more present than in a UAV system. Therefore, a major portion of any UAV HSI evaluation must be dedicated to identifying the degree of automation present in the system and the effects of that automation on overall system performance. The degree of automation determines the level of control the operator has over that UAV system. Levels of human-system control can range from total, manual control to supervisory control to a level Reising has termed 'co-operative functioning' [52].

In manual control the human specifies the goals and functions to be accomplished and the machine carries out the tasks [52]. Aircraft systems under manual control require the ground station-based operator to fly the UAV using traditional aircraft controls such as a stick and throttle (e.g., Predator). UAV operators may use aircraft-mounted camera video to obtain the visual information required to successfully launch, fly and land the

aircraft. Other aircraft subsystem health, navigation, armament or sensor information can be displayed on one or more visual displays.

In supervisory control the human still specifies the goals but the machine carries out both the tasks and functions [52]. Aircraft systems under supervisory control are typically operated via a mouse and/or keyboard with the pilot entering macro level commands that guide the aircraft through a series of flight operations. Supervisory control tasking can range from simply entering a new aircraft airspeed to changing course to selecting a whole new mission. All of these task change inputs are typically executed via the keyboard or the mouse.

In systems exhibiting a 'cooperative functioning control' architecture the human and machine interact at all levels and either of them can perform the goals, functions and tasks [52]. Although the human will always maintain ultimate control over the whole system, the human and machine operate more as partners on a team by dynamically sharing all levels of control authority. Aircraft systems operating at a true cooperative functioning control level have not been developed yet. Unmanned aircraft systems, such as Predator, have also been developed that operate at a level approaching cooperative functioning. In these systems the aircraft may send imagery back to a human-operated ground control station for human review. The human then designates desired targets and sends this information back to the aircraft for target destruction command execution. In this situation, however, the aircraft does not have the authority to change the types of targets designated or to decide whether to attack the targets at all.

In flight test the level of pilot-UAV control must be identified and the effects of that control level on pilot-UAV system performance must be assessed for specific phases of the intended mission. As described earlier, increased automation does not necessarily result in reduced workload or increased pilot-UAV system performance. In addition to automation-induced complacency, system malfunction error recovery can cause significant problems for the UAV pilot. UAV pilots may have much more difficulty resolving unexpected problems due to being 'out-of-the-loop' when problems occur. In this instance, workload can be very high while the pilot attempts to ascertain the nature and degree of the system malfunction (reduced situation awareness) and decide on an appropriate course of action to recover from the malfunction. The opportunity to directly measure UAV pilot situation awareness and workload during or after these types of system malfunctions does not always present itself during the course of flight test. However, if the opportunity does arise, a thorough error analysis and the events and conditions leading up to that error can prove quite beneficial.

4.2 Unique UAV Flight Test Opportunities

4.2.1 Operator Measures of Performance

Due to required pilot protective equipment and limited cockpit space and resources, there are some measures of performance that cannot be used in manned aircraft cockpits but can be used in evaluations of UAV workstations. Some examples include expert observations, audio and video recordings, and eye/head-tracking measures. Expert observations can be made of operators performing various tasks during flight test, noting the level of situation awareness and amount of operator workload. Audio and video recordings can be made of UAV operators while they are conducting specific operations for subsequent analyses of operator behaviors, such as the number of button pushes or operator actions to accomplish a specific task.

An eye/head-tracking system can be used to determine which display and control objects the UAV operator is looking at and for how long. This information can be used to determine display symbol appearance, disappearance or status change detection rates which can then be used to compute situation awareness metrics. Display object dwell times can be used to derive estimates of text readability, object visibility or information

complexity and support task timeline analyses. Eye gaze patterns can be used to determine what information the UAV operator is using to make specific decisions and the speed and order in which s/he obtains that information. Objective HSI flight test data is difficult to obtain, however, the UAV operator workstation environment lends itself to such opportunities.

4.2.2 Simulation Fidelity

Modeling and simulation of UAV operator workstations is easier to accomplish and has higher fidelity than manned aircraft simulators. Simulation of a manned cockpit containing realistic out-the-window displays, pilot controls and displays and visualization and aircraft handling software requires large amounts of money, facility space, hardware, software, and quality support personnel. Most UAV workstations can and do serve as their own simulator. All the simulator workstation controls, displays, display symbology and controls functionality are exactly the same as the real workstation. This allows evaluators to use measures of performance normally prohibited in manned aircraft cockpits, such as SAGAT, to collect performance data on actual UAV workstations under near-real mission conditions. Human-system integration data can be collected from such UAV simulators at a much reduced cost. And the data can be more easily generalized to the actual UAV workstation than its manned aircraft counterpart.

5.0 CONCLUSIONS

The measures used to conduct HSI evaluations of manned aircraft are not considerably different than those used to evaluate unmanned systems. The differences lie in the evaluation areas that are focused upon and the emphasis they receive. Situation awareness and workload are always important factors. They are paramount in assessing the adequacy of a UAV system's human interface. One new evaluation area is level of operator control. UAV-operator system control levels must be identified and the human interface evaluated to verify that the interface adequately supports the associated level of control. Finally, there are unique and valuable data collection opportunities provided by UAV control stations versus. those of manned cockpits.

6.0 REFERENCES

[1] Roscoe, A.H. In-flight assessment of workload using pilot ratings and heart rate. In A.H. Roscoe (Ed.), The practical assessment of pilot workload AGARDograph No. 282 (pp. 78-82) 1987a. Neuilly Sur Seine, France: AGARD.

[2] Spravka, J.J. *Global Hawk Ground Control Station Human Factors Characterization, AFFTC-TR-99-39.* Air Force Flight Test Center, Edwards AFB, CA., 2000.

[3] Geddes, N.D., Brown, J.L., Andes, R.C., Hoskin, N.J., Lee, B., Coombs, C., and Lafferty, L. *Intelligent Control for Automated Vehicles: A Decision Aiding Method for Coordination of Multiple Uninhabited Tactical Aircraft.* Applied Systems Intelligence, Inc. ARPA Order 0611. U.S. Army Missile Command Contract No. DAAH 01-96-C-R245, 1997.

[4] Dixon, S.R., Wickens, C.D. and Chang, D. Comparing Quantitative Model Predictions to Experimental Data in Multiple-UAVs Flight Control. *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting*, pp.104 – 108, 2003. Denver, CO: Human Factors and Ergonomics Society.

[5] Calhoun, G., Draper, M., Ruff, H., Fontejon', J. and Guilfoos, B. "Evaluation of Tactile Alerts For Control Station Operation." *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting*, pp.2118 – 2122, 2003. Denver, CO: Human Factors and Ergonomics Society.

[6] Draper, M., Calhoun, G., Ruff, H., Williamson, D. and Barry, T. "Manual Versus Speech Input for Unmanned Aerial Vehicle Control Station Operations." *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting*, pp.109 – 113, 2003. Denver, CO: Human Factors and Ergonomics Society.

[7] Bell, B.G. and Cooke, N.J. "Cognitive Ability Correlates of Performance on a Team Task." *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting*, pp.1087 – 1091, 2003. Denver, CO: Human Factors and Ergonomics Society.

[8] van Breda, L. *Operator Performance in Multi-Maritime Unmanned Air Vehicle Control*, TNO Human Factors Research Institute, Report #TNO-TM 1995 A-76, 1995.

[9] Chadwick, R. A., Gillan, D. J., Simon, D. and Pazuchanics, S. Cognitive Analysis Methods for Control of Multiple Robots: Robotics on $5 a Day. Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting, 2004.

[10] Wierwille, W., & Casali, J. "A validated rating scale for global mental workload measurement applications." *Proceedings of the Human Factors Society 27th Annual Meeting,* pp. 129-133, 1983.

[11] Wickens, C.D. *Engineering Psychology and Human Performance*. Columbus, OH: Merrill, 1984.

[12] Boer, L.C. "Taskomat: Beteknis van" Report IZF, 1993 A-29. Soesterberg, NL: TNO Human Factors Research Institute.

[13] Veltman, J.A. and Gaillard, A.W.K. "Evaluation of subjective and psychological measurement techniques for pilot workload." Report IZF, 1993 A-5, Soesterberg, NL: TNO Human Factors Research Institute.

[14] Breda van, L. & Passenier, P.O. "An exploratory study of the human-machine interface for controlling maritime unmanned air vehicles." Report IZF, 1993 A-10, Soesterberg, NL: TNO Human Factors Research Institute.

[15] Endsley, M.R. "Direct measurement of situation awareness: validity and use of SAGAT." *Situation Awareness and Measurement.* Endsley, M.R. and Garland D.J (Eds). Mahway, NJ: Lawrence Erlbaum Associates.

[16] Parasuraman, R. "Effects of adaptive function allocation on human performance." In D.J. Garland and J.A. Wise (Eds.), <u>Human Factors and Advanced Aviation Technologies</u> (147-157), 1993. Daytona, FL: Embry-Riddle Aeronautical University Press.

[17] Parasuraman, R. & Mouloua, M. <u>Automation and human performance: Theory and applications</u>. Mahway, NJ:Erlbaum.

[18] Parasuraman, R., Mouloua, M. & Hilburn, B. "Adaptive aiding and adaptive task allocation enhance human-machine interaction." In M. Scerbo and M. Mouloua (Eds.), Automation technology and human performance: Current research and trends (pp. 119-123). Mahwah, NJ:Erlbaum.

[19] Parasuraman, R. & Riley, V. "Humans and automation: Use, misuse, disuse, abuse." Human Factors, 39, 230-253, 1997.

[20] Weiner, E.L. "Cockpit automation." In E.L. Wiener and D.C. Nagel, (Eds.), Human Factors in Aviation. (pp. 433-461). San Diego, CA: Academic Press.

[21] Woods, D.D. "Decomposing automation: Apparent simplicity, real complexity." In R. Parasuraman and M. Mouloua (Eds.) Automation and Human Performance: Theory and Applications (pp. 1-17). Mahwah, NJ: Erlbaum.

[22] Sarter, N.B. & Woods, D.D. "How in the world did we ever get into that mode? Mode error and awareness in supervisory control." Human Factors, 37, 5-19, 1995.

[23] National Transportation Safety Board Eastern Airlines L-1011, Miami, Florida, December 29, 1972 (Report No. NTSB-AAR-73-14). Washington, DC.

[24] National Transportation Safety Board China Airlines Boeing 747-SP, N4522V, 300 nautical miles northwest of San Francisco, California 1985 (Report No. NTSB-AAR-86-03). Washington, DC.

[25] Stein, K.J. "Human factors analyzed in 007 navigation error." Aviation Week & Space Technology, 165-167, 1983.

[26] Kessel, C.J. & Wickens, C.D. "The transfer of failure-detection skills between monitoring and controlling dynamics." Human Factors, 24, 49-60, 1982.

[27] Knapp, R.K. & Vardman, J.J. "Response to an automated function failure cue: an operational measure of complacency." In Proceedings of the Human Factors Society, 35th Annual Conference, 1991. San Francisco, CA: Human Factors Society.

[28] Parasuraman, R., Molloy, R. & Singh, I.L. "Performance consequences of automation induced "complacency"." International Journal of Aviation Psychology, 3, 1-23, 1993.

[29] Parasuraman, R., Sheridan, & T.B, Wickens, C.D. "A model for types and levels of human interaction with automation." IEEE Transactions on Systems, Man, and Cyberntics-Part A: Systems and Humans, 30, 286-297, 2000.

[30] Woods, D. D., Sarter, N. B., and Billings, C. E. "Automation surprises." In G. Salvendy (Ed.), Handbook of Human Factors/Ergonomics, (2nd ed.) New York, NY: Wiley.

[31] Lee, J.D., & Moray, N. "Trust, control strategies and allocation of function in human-machine systems." Ergonomics, 35, 1243-1270, 1992.

[32] Lee, J.D., & Moray, N. "Trust, self-confidence, and operators' adaptation to automation." International Journal of Human-Computer Studies, 40, 153-184, 1994.

[33] Masalonis, A.J. Effects of situation-specific reliability on trust and usage of automated decision aids. Unpublished doctoral dissertation, 2000. Washington, DC: The Catholic University of America.

[34] Muir, B.M. "Trust between humans and machines, and the design of decision aides." In E. Hollnagel, G. Mancini, and D.D. Woods (Eds.), Cognitive engineering in complex dynamic worlds (pp. 71-83). London: Academic Press.

[35] Sheridan, T.B. "Trustworthiness of command and control systems." In Proceedings of the international Federation of Automatic Control Conference on Man-Machine Systems (pp. 427-431, 1988). Elmsford, NY: Pergamon.

[36] Wickens, C.C. "Designing for situation awareness and trust in automation." In Proceedings of the IFAC Conference on Integrated Systems Engineering, 1994. Baden-Baden, Germany: International Federation of Automatic Control.

[37] Hurst, K., & Hurst, L. Pilot error: The human factors. New York: Aronson, 1982.

[38] Sumwalt, R.L., Morrison, R., Watson A. & Taube, E. "What ASRA data tell about inadequate flight crew monitoring." In Proceedings of the 9th International Symposium on Aviation Psychology (pp. 977-982, 1997). Columbus, OH: Ohio State University.

[39] Thackray, R.I. and Touchstone, R.M. "Detection efficiency on air traffic control monitoring task with and without computer aiding." Aviation, Space and Environmental Medicine, 60, 744-748, 1989.

[40] Parasuraman, R., Mouloua, M. & Molloy, R. "Monitoring automation failures in human-machine systems." In M. Mouloua & R. Parasuraman (Eds.), Human performance in automated systems: Current research and trends (pp. 45-49). Hillsdale, NJ: Lawrence Erlbaum Associates.

[41] Molloy, R. & Parasuraman, R. "Monitoring an automated system for a single failure; Vigilance and task complexity effects." Human Factors, 38(2), 311-322, 1996.

[42] Singh, I.L. Molloy, R. & Parasuraman, R. "Automation-induced monitoring inefficiency: role of display location." International Journal of Human-Computer Studies (46), 17-30, 1997.

[43] Duley, J.A., Westerman, S. Molloy, R. & Parasuraman, R. "Effects of display superimposition on monitoring of automation." In Proceedings of the 9th International Symposium on Aviation Psychology (pp. 322-328, 1997). Columbus, OH: Ohio State University.

[44] Baddely, A.D. Working memory. Oxford, U.K.

[45] Crocoll, W.M., & Coury, B.G. "Status or recommendation: Selecting the type of information for decision aiding." In Proceedings of the Human Factors and Ergonomics Society 34th Annual Meeting (pp. 1524 – 1528, 1990). Santa Monica, CA: Human Factors and Ergonomics Society.

[46] Sarter, N. and Schroeder, B.K. "Supporting a decision-making and action selection under time pressure and uncertainty: The case of the in-flight icing." Human Factors, 43(4), 573-583, 2001.

[47] Rovira, E., McGarry, K. & Parasuraman, R. "Effects of unreliable automation on decision making in command and control." In Proceedings of the 46th Annual Meeting of the Human Factors and Ergonomics Society (pp. 428-432, 2002). Santa Monica, CA: Human Factors and Ergonomics Society.

[48] Rovira, E., Zinni, M. & Parasuraman, R. "Effects of information and decision automation on multi-task performance." In Proceedings of the 46th Annual Meeting of the Human Factors and Ergonomics Society (pp. 327-331, 2002). Santa Monica, CA: Human Factors and Ergonomics Society.

[49] Galster, S.M., Bolia, R.S., Roe, M.M., & Parasuraman, R. "Effects of automated cueing and decision implementation in a visual search task." In Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting (pp. 321-325, 2001). Santa Monica, CA: Human Factors and Ergonomics Society.

[50] Galster, S.M., Bolia, R.S., & Parasuraman, R. "Effects of information automation and decision-aiding cueing on action implementation in a visual search task." In Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting (pp. 438-442, 2002). Santa Monica, CA: Human Factors and Ergonomics Society.

[51] Galster. S.M. "An Examination of Complex Human-Machine System Performance Under Multiple Levels and Stages of Automation." Crew System Interface Division AFRL/HEC Bldg 248, 2255 H Street Wright-Patterson AFB, OH 45433-7022 USA, AFRL-HE-WP-TR-2003-0149.

[52] Reising, J. "Uninhabited Military Vehicles: What is the Role of the Operators?" Crew System Interface Division AFRL/HEC Bldg 248, 2255 H Street Wright-Patterson AFB, Oh 45433-7022 USA, RTO-MP-088, 2003.

7.0 ACRONYM LIST

AFFTC – Air Force Flight Test Center

ASRS - Aviation Safety Reporting System

CMT – Continuous memory task

GPA – Grade Point Average

HSI – Human-System Integration

NASA – National Aeronautics and Space Administration

MUAV – Maritime unmanned air vehicle

RMS - Root-mean-squared

SAGAT – Situation Awareness Global Assessment Technique

SWAT – Subjective Workload Assessment Technique

TLX – Time Load Index

UAV – Unmanned Air Vehicle

USAF – SAM – Unites States Air Force – School of Aerospace Medicine

UTA – Uninhabited Tactical Aircraft